图书情报工作
第63卷 第5期 2019年3月

我国省级公共图书馆评价研究

——以大众点评网数据为例

■ 张文亮　付枭男　王彦芳
辽宁师范大学政府管理学院　大连 116029

摘要：[目的/意义]探究我国省级公共图书馆的公众评价现状，比较华东、华中、华南、华北、西北、西南、东北共7个地区31个省内39个省级公共图书馆的建设差异，为公共图书馆未来建设方向提供建议。[方法/过程]通过《中国公共图书馆事业发展基础数据概览》《中国图书馆年鉴2016》和中国人民共和国国家统计局网站获取公共图书馆客观评估数据，确定数据获取的几个维度，应用python网络爬虫技术获取我国省级公共图书馆在大众点评上的赋值评分和用户评论，应用频数分析、偏度分析、相关性分析和情感分析方法从读者视角分析我国公共图书馆评价各要素之间的关系，提出建设性建议。[结果/结论]通过数据分析得出研究结论：①均衡差异，涉及地区差异和评价内容差异；②增强对用户的吸引力，尤其是偏远地区；③关注用户情感倾向，重点关注读者关注的热点词，同时要注意消除负面情感带来的影响。

关键词：图书馆评价　Python　评价分析　大数据
分类号：G251
DOI:10.13266/j.issn.0252-3116.2019.05.006

随着我国公共图书馆事业的不断发展和公共图书馆服务水平的不断提高，公共图书馆各项服务工作正在逐步进入科学化、标准化的管理阶段。为提高公共图书馆的综合服务能力，公共图书馆评价的研究成为公共图书馆关注的焦点问题。

我国公共图书馆评价的研究起源于20世纪80年代，最早的相关研究是在借鉴、比较中进行的，宋运郊[1]、施士宇[2]等学者试图从国外公共图书馆评价方法中为百废待兴的公共图书馆事业提供指引；1986年，随着文化部下达的有关公共图书馆评估工作的指示，为满足图书馆相关建设需要，以县级公共图书馆评估研究为起点[3]，以评估研究总体要求为主要建设依据[4]，全国各地区、各等级公共图书馆的评估评价相关讨论呈现出强劲的研究态势[5-6]；理论研究的加深促成了相关探索的细化，20世纪90年代，公共图书馆总体评价理论的探讨转到了图书馆评估指标体系建设[7-9]、从业人员工作评价[10]、藏书评价[11-13]、服务工作评价[14-15]、服务效果评价[16-17]等多角度的研究范畴之中；21世纪初期，图书馆定量化、电子化、智能化的大潮席卷而来，公共图书馆评价研究也卷入了技术方法研究的思潮[18-21]，实现了理论与技术的结合统一；近10年来，得利于"十二五""十三五"规划纲要中对于公共图书馆服务的重视[22]，以及公共图书馆评估工作的顺利展开[23-25]，对于公共图书馆服务评价[26-27]、绩效评估[28-29]、评估工作[30]的理论研究又重新占据主流地位。

从实践层面来看，我国公共图书馆评估工作经历六届，历时23年，结合专家评估，从定量和定性的角度出发，以馆舍面积、馆藏数量、经费、技术设备、读者满意度、重点文化工程、现代化技术条件、数字资源服务以及本区域服务体系规划与共建共享等为关键指标[31]。这种评估是基于专家视角的评价，对读者的评价仅从读者满意度一个方面考察。因此，本研究从读者视角出发，通过对大众点评网中用户到馆实际体验的评价信息为研究对象，探讨读者视角的公共图书馆评价。

1　数据来源与研究工具

1.1　数据来源

大众点评网于2003年成立，是中国领先的集本地

作者简介：张文亮（ORCID：0000-0002-4606-3686），副教授，博士，E-mail：zwl2002@sina.com；付枭男（ORCID：0000-0003-3264-2289），硕士研究生；王彦芳（ORCID：0000-0002-2289-8443），硕士研究生。

收稿日期：2018-06-20　修回日期：2018-09-16　本文起止页码：51-60　本文责任编辑：王传清

生活与交易信息于一身的网络平台，也是全球最早建立的独立第三方消费点评网站。它具有为用户提供商户信息、自身及其他消费者点评及消费优惠等信息服务的功能。它首创并领导的第三方评论模式已成为互联网的一个新热点，用户可以自由发表对商家的评论及自身的感受心得。几乎所有的评论信息都来源于大众，最终也为大众提供参照[32]。

本研究数据来源于大众点评网，数据采集的对象是我国华东、华中、华南、华北、西北、西南、东北共7个地区的31个省（除港澳台）内的39个省级公共图书馆，其中青海省图书馆未被大众点评网站收录，天津、广东、广西、福建、上海、湖南、河南、重庆这几个省及直辖市都收录多个省级公共图书馆。实际数据爬取日期截至2018年3月12日。

在大众点评网上按照《中国图书馆年鉴2016》中收录的省级公共图书馆的名称进行逐一检索，确定了被大众点评收录并且有评价星级的39个省级公共图书馆（包含分馆），见表1，并将其在大众点评网对应的网址以Json（JavaScript Object Notation）数据格式保存。Json是基于JavaScript（Standard ECMA－262 3rd Edition - December 1999）的一个子集，是一种轻量级的数据交换格式，易于阅读和编写，同时也易于机器进行解析和生成[33]。

表1 省级公共图书馆地区数量分布统计

地区	省份	馆名	数量（个）	地区	省份	馆名	数量（个）
华东地区	山东	山东省图书馆	1	华北地区	北京	首都图书馆	1
	江苏	南京图书馆	1		天津	天津图书馆（复康路分馆）	3
	安徽	安徽省图书馆	1			少年儿童图书馆	
	浙江	浙江图书馆	1			天津图书馆（文化中心总馆）	
	福建	福建省图书馆	2		河北	河北省图书馆	1
		福建省少年儿童图书馆			山西	山西省图书馆	1
	上海	上海少年儿童图书馆	2		内蒙古	内蒙古图书馆	1
		上海图书馆		西北地区	宁夏	宁夏图书馆	1
华南地区	广东	广东省立中山图书馆（文明路总馆）	2		新疆	新疆自治区图书馆	1
		广州图书馆			青海	青海图书馆	未收录
	广西	广西壮族自治区图书馆	2		陕西	陕西省图书馆	1
		南宁市图书馆			甘肃	甘肃省图书馆	1
	海南	海南省图书馆	1	西南地区	四川	四川省图书馆	1
华中地区	湖北	湖北省图书馆（新馆）	1		云南	云南省图书馆	1
	湖南	湖南省少年儿童图书馆	2		贵州	贵州省图书馆	1
		湖南图书馆			重庆	重庆少年儿童图书馆	2
	河南	河南省儿童图书馆	2			重庆市图书馆	
		河南图书馆			西藏	西藏图书馆	1
				东北地区	辽宁	辽宁省图书馆	1
	江西	江西省图书馆	1		吉林	吉林省图书馆	1
					黑龙江	黑龙江省图书馆	1

1.2 研究工具

研究使用Pandas、Zipfile、Numpy、Matplotlib等多个Python中有关数据分析和挖掘相关的库。由于数据分析对象是我国省级公共图书馆，数据量大，数据整体压缩后才进行数据上传，因此引入Zipfile库，对数据进行解压处理。Pandas纳入了大量库和一些标准的数据模型，提供高效地操作大型数据集所需的工具，以及大量能快速便捷地处理数据的函数和方法。Numpy库可以提供许多高级的数值编程工具，如矩阵数据类型、矢量处理以及精密的运算库，专门用于进行严格的数字处理。Matplotlib是一个Python的2D绘图库，仅需要几行代码，便可以通过Matplotlib生成绘图，如直方图、条形图、散点图等，能够使得数据结果更加直观得展示出来。本研究通过引入这些数据分析库，对获取到的数据进行评分对比、描述、相关性等分析，从而发现不足，提出应对策略。

2 数据分析

本研究应用网络爬虫技术爬取大众点评网评价的所有信息，包括省级公共图书馆的名称、星级评价、书

籍评分、环境评分、服务评分、评论内容、评论内容数量分布(包括好评数、中评数、差评数和带图片的评论数)、人均价格等数据,见表2。

获取到的部分图书馆评价信息见表3。

表2 大众点评网数据项及其解释说明

数据项	数据解释
图书馆名称(libName)	对照《中国图书馆年鉴2016》和大众点评网收录名称确定可检索到的省级公共图书馆名称
星级评价(libCommentsRanks)	大众点评网根据用户评价官方赋予的星级
书籍评分(bookReview)	大众点评网根据用户对书籍的整体评价官方赋予的评分
环境评分(envReview)	大众点评网根据用户对环境的整体评价官方赋予的评分
服务评分(servReview)	大众点评网根据用户对服务的整体评价官方赋予的评分
评论内容(libCommm)	大众点评网用户评论(本研究只对文本内容进行分析,对图片不进行分析)
评论内容数量分布(reviewCount)	大众点评网站官方对评论内容进行了好中差以及带有图片评论的划分和数量的统计
人均价格(avgPrice)	大众点评网根据用户消费情况统计给出的平均价格(可能包含周边增值服务)

表3 部分图书馆基本信息

图书馆名称	人均价格(元)	书籍评分	环境评分	服务评分	评论内容的数量分布
内蒙古图书馆	0	7.5	7.5	7.3	{‘好评’:2,‘差评’:0,‘图片’:1,‘中评’:0}
宁夏图书馆	143	7.9	8.2	7.5	{‘好评’:35,‘差评’:5,‘图片’:13,‘中评’:3}
山东省图书馆	67	7.6	7.6	7.3	{‘好评’:58,‘差评’:5,‘图片’:19,‘中评’:18}
上海少年儿童图书馆	91	9.2	9.1	8.4	{‘好评’:375,‘差评’:14,‘图片’:118,‘中评’:39}
上海图书馆	64	8.5	8.4	7.9	{‘好评’:1934,‘差评’:152,‘图片’:639,‘中评’:397}
山西省图书馆	0	8.6	8.6	8.3	{‘好评’:50,‘差评’:2,‘图片’:31,‘中评’:2}
陕西省图书馆	83	9.1	9.0	8.4	{‘好评’:215,‘差评’:4,‘图片’:87,‘中评’:27}
首都图书馆	45	9.1	9.2	8.8	{‘好评’:670,‘差评’:25,‘图片’:238,‘中评’:85}
四川省图书馆	99	9.0	9.1	8.6	{‘好评’:212,‘差评’:24,‘图片’:195,‘中评’:30}
天津少年儿童图书馆	0	7.5	7.5	7.3	{‘好评’:17,‘差评’:3,‘图片’:9,‘中评’:4}
天津图书馆(复康路分馆)	45	9.0	8.6	8.3	{‘好评’:399,‘差评’:16,‘图片’:58,‘中评’:61}
天津图书馆(文化中心分馆)	78	9.2	9.3	8.8	{‘好评’:435,‘差评’:45,‘图片’:215,‘中评’:52}
新疆自治区图书馆	0	7.1	7.1	7.1	{‘好评’:2,‘差评’:0,‘图片’:0,‘中评’:2}
西藏图书馆	0	7.4	7.3	7.3	{‘好评’:4,‘差评’:0,‘图片’:2,‘中评’:0}
云南省图书馆	171	7.9	7.8	7.6	{‘好评’:43,‘差评’:1,‘图片’:13,‘中评’:12}
浙江图书馆	42	9.1	9.0	8.7	{‘好评’:198,‘差评’:3,‘图片’:47,‘中评’:16}
广东省立中山图书馆(文明路总馆)	40	8.6	8.6	8.0	{‘好评’:254,‘差评’:8,‘图片’:123,‘中评’:42}
安徽省图书馆	0	7.9	7.8	7.9	{‘好评’:27,‘差评’:1,‘图片’:15,‘中评’:6}
重庆少年儿童图书馆	0	7.7	7.7	7.7	{‘好评’:2,‘差评’:0,‘图片’:2,‘中评’:0}
重庆市图书馆	62	8.8	9.1	8.7	{‘好评’:55,‘差评’:0,‘图片’:29,‘中评’:10}
福建省少年儿童图书馆	0	7.8	8.1	7.8	{‘好评’:21,‘差评’:1,‘图片’:21,‘中评’:6}
福建省图书馆	0	8.0	7.8	7.5	{‘好评’:33,‘差评’:3,‘图片’:29,‘中评’:6}
甘肃省图书馆	0	7.4	7.5	7.3	{‘好评’:7,‘差评’:0,‘图片’:6,‘中评’:2}
广西壮族自治区图书馆	0	7.7	7.7	7.2	{‘好评’:30,‘差评’:1,‘图片’:12,‘中评’:4}
广州图书馆	13	9.2	9.2	9.0	{‘好评’:581,‘差评’:34,‘图片’:398,‘中评’:74}
贵州省图书馆	0	7.0	7.4	7.1	{‘好评’:6,‘差评’:1,‘图片’:5,‘中评’:0}
海南省图书馆	0	7.4	7.5	7.3	{‘好评’:27,‘差评’:2,‘图片’:14,‘中评’:7}
河北省图书馆	0	8.4	8.5	7.8	{‘好评’:47,‘差评’:5,‘图片’:21,‘中评’:6}
河南省儿童图书馆	0	7.6	7.5	7.5	{‘好评’:13,‘差评’:0,‘图片’:10,‘中评’:2}
河南省图书馆	0	7.4	7.4	7.3	{‘好评’:20,‘差评’:2,‘图片’:6,‘中评’:11}
湖北省图书馆新馆	32	9.3	9.3	9.2	{‘好评’:274,‘差评’:8,‘图片’:155,‘中评’:21}
湖南省少年儿童图书馆	0	6.8	6.7	6.8	{‘好评’:1,‘差评’:1,‘图片’:1,‘中评’:1}
湖南图书馆	0	8.1	8.0	8.0	{‘好评’:69,‘差评’:2,‘图片’:28,‘中评’:10}
江西省图书馆	43	7.5	7.4	7.4	{‘好评’:34,‘差评’:2,‘图片’:24,‘中评’:2}
吉林省图书馆	25	9.1	9.2	9.0	{‘好评’:116,‘差评’:8,‘图片’:67,‘中评’:4}
辽宁省图书馆	39	9.1	9.2	9.0	{‘好评’:242,‘差评’:20,‘图片’:82,‘中评’:58}
南京图书馆	108	9.1	9.1	8.7	{‘好评’:769,‘差评’:16,‘图片’:231,‘中评’:50}
南宁市图书馆	0	7.3	7.3	7.3	{‘好评’:6,‘差评’:0,‘图片’:5,‘中评’:0}
黑龙江省图书馆	71	9.1	9.1	8.7	{‘好评’:84,‘差评’:6,‘图片’:42,‘中评’:13}

随后查看数据项(数值型数据和文本数据)。由于网络原因,在数据抓取的过程中存在数据缺失现象,如人均价格,正常情况下为39条数据,但存在3条数据缺失,利用fillna()对数据进行补全。

为了进一步对数据分布情况有整体性把握,应用dataframe describe()查看数值型数据的具体分布情况,并对"评论内容数量分布"进行拆分,将好评数、差评数、中评数单独成列,方便观察和分析数据,如表4所示:

表4　数值型信息概览

	人均价格(元)	书籍评分	环境评分	评论总数	服务评分	好评数	中评数	差评数
总数	39.000 000	39.000 000	39.000 000	39.000 000	39.000 000	39.000 000	39.000 000	39.000 000
平均值	34.897 436	8.205 128	8.215 385	258.256 410	7.943 590	188.897 436	27.769 231	10.769 231
标准差	44.421 898	0.779 996	0.781 206	524.237 879	0.678 541	347.534 458	64.759 662	25.362 074
最小值	0.000 000	6.800 000	6.700 000	2.000 000	6.800 000	1.000 000	0.000 000	0.000 000
下四分位	0.000 000	7.500 000	7.500 000	26.000 000	7.300 000	18.500 000	2.000 000	1.000 000
中位	13.000 000	8.000 000	8.100 000	61.000 000	7.800 000	47.000 000	7.000 000	3.000 000
上四分位	63.000 000	9.100 000	9.100 000	305.000 000	8.650 000	228.000 000	28.500 000	8.000 000
最大值	171.000 000	9.300 000	9.300 000	3 044.000 000	9.200 000	1 934.000 000	397.000 000	152.000 000

数据来源:本研究整理

为进一步探究所获取数据间的关系及呈现规律,本研究分别应用频数分析、偏度分析、相关性分析以及情感倾向分析:①频数分析:频数也称"次数",对总数据按某种标准进行分组,统计出各个组内含个体的个数。②偏度分析:偏度是对样本数据构成的分布的对称性状况的描述,可以用来衡量数据集的分布形状。它取值通常在(-3,3)之间,衡量数据集的对称程度[34]。偏度系数越接近0,说明数据集越对称,越远离0则表明数据集越不对称,偏度系数值越大数据越分散。如果偏度系数为正,则表明数据在右边更为分散,若为负,则说明数据在左侧更为分散。③相关性分析:指对两个或多个具备相关性的变量元素进行分析,从而衡量两个变量因素的相关密切程度。相关性的元素之间需要存在一定的联系或者概率才可以进行相关性分析。④情感倾向分析:情感倾向分析就是从用户评论的文本内容中识别出观点,分析其感情倾向,并抽取相关的观点要素。

2.1　频数分析

据本研究统计,39个省级公共图书馆在大众点评网上主要呈现6个星级,其中,五星商户有10个,准五星商户7个,四星商户9个,准四星商户12个,三星商户1个,二星商户0个,具体如表5所示:

表5　图书馆星级分类情况

图书馆星级	数量(个)	图书馆名称
五星商户	10	上海少年儿童图书馆、陕西省图书馆、首都图书馆、四川省图书馆、天津图书馆(文化中心分馆)、浙江图书馆、广州图书馆、湖北省图书馆新馆、吉林省图书馆、南京图书馆
准五星商户	7	上海图书馆、山西省图书馆、天津图书馆(康复路分馆)、广东省立中山图书馆(文明路总馆)、重庆市图书馆、辽宁省图书馆、黑龙江省图书馆
四星商户	9	宁夏图书馆、山东省图书馆、云南省图书馆、安徽省图书馆、重庆少年儿童图书馆、福建省少年儿童图书馆、福建省图书馆、河北省图书馆、湖南图书馆
准四星商户	12	内蒙古图书馆、天津少年儿童图书馆、新疆自治区图书馆、西藏图书馆、甘肃省图书馆、广西壮族自治区图书馆、贵州省图书馆、海南省图书馆、河南省儿童图书馆、河南省图书馆、江西省图书馆、南宁市图书馆
三星商户	1	湖南省少年儿童图书馆
二星商户	0	-

研究发现,星级很大程度上决定于到馆阅读人群对该馆的书籍评分、服务评分和环境评分,因此,针对以上3个因素进行数值分析,以便对这39个图书馆评分值范围进行大致描述,见图1(横坐标是三个因素评分的数值范围,纵坐标是对应的各评分数的频数值)。

由图1可见,3类评分值分布的范围相对宽泛,但每组值的分布频数存在差异,比如书籍评分和环境评分数值最多的都位于(9.0,9.5)这一区间,而服务评分众数区间在(7.0,7.5)区间,一定程度上反应出我国省级公共馆在馆藏和阅读环境方面被到馆读者认可程度较高,但服务方面还存在一些欠缺,有待加强。

2.2　评价因素偏度分析

在频数分析的基础上,本研究进一步进行偏度分析(见图2),以查看数据分布特征规律。

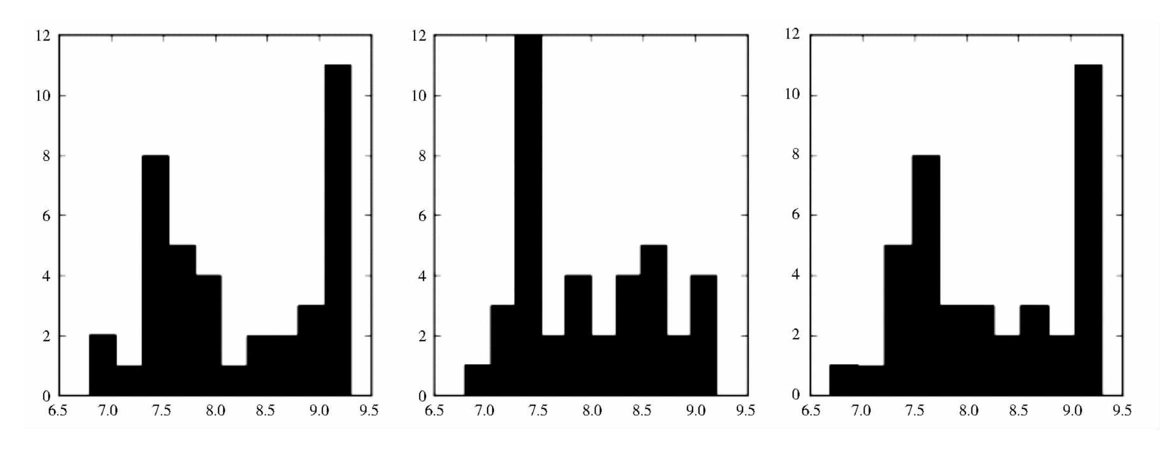


图1　书籍评分、服务评分、环境评分频数分布

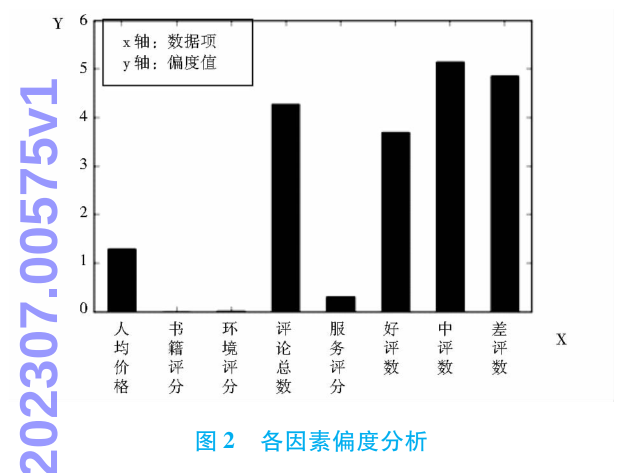


图2　各因素偏度分析

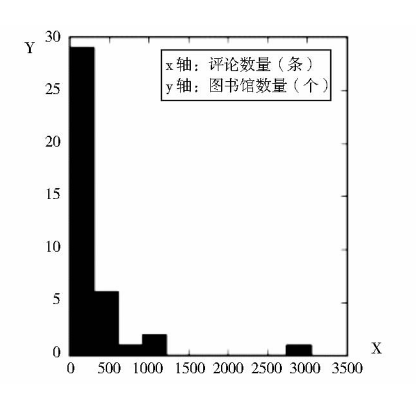


图3　评论数分布直方图

从偏度分析图中可知,书籍评分、服务评分、环境评分3类数值分布比较集中,偏度值基本都在0附近,而且不存在负偏,说明书籍、服务和环境三者数据集高度对称,可见大众点评网将公共图书馆划分为这3个维度赋分是具有科学意义的;人均价格偏度在(1,2)之间,在合理范围之内。本研究考察的省级公共图书馆是免费开放和使用的,这里评论中大众点评网标注的人均价格并不是指图书馆入门等价格,而是图书馆周边服务或增值服务的收费情况,例如周边服务如餐厅、停车场收费等,增值服务如复印书籍稿件等;通过图2不难发现,各个图书馆之间的评论数存在较大差异,好评数、差评数、中评数都属于评论数的一部分,三者的偏度值基本上等同于评论数偏度值,偏度值大都在(4,5)之间,超出(-3,3)范围,说明评论数数据分散度比较高。数值比较后存在极大值或极小值,一定程度上说明到馆并应用点评网进行评论的人数和条数存在较大差异,为此本研究单独对评论数进行分析,绘制评论分布直方图见图3。

从图3中可以直观看出,数值大都集中在(0,500)这一区间,图书馆数量大约占74%左右,并且存在一个极大值,即上海图书馆的3 044条评论数,由此可见图书馆评论数偏度分析差异很大。各省级图书馆评论数量见表6。

2.3　评价因素相关性分析

为确定各个因素之间的相关性,本研究利用Pathon的dataframe.corr()函数给出各个因素间的相关性数值(见表7),值越大,相关性与强。

由表7可知,强相关性的前3位是书籍评分与环境评分0.98,中评数与差评数相关性0.975,服务评分与环境评分是0.96;处于弱相关性的末3位是人均价格与中评数0.22,中评数与服务评分相关性0.22,人均价格与差评数0.23。

为更方便直观地看各个因素之间的关系,以热点图的方式呈现出来,颜色越深,相关性越大。由图4可知,书籍评分、环境评分、服务评分三者之间联系最为紧密;人均价格与其他因素之间相关系数值小;好评数、中评数、差评数三者之间相关性也比较靠前。结合图4可知,书籍评分跟环境评分相关性最大,占0.98,其次是服务评分,为0.96。环境评分与书籍评分相关性最大,高达0.98,其次是服务评分,占0.96。服务评分与环境评分的相关性也最大,占0.96,其次是与书

籍评分,占0.96。由此可见,这三者之间的相关性很密切,是大众点评网衡量公共图书馆的重要指标。

表6 各省级图书馆评论数量

名次	图书馆名称	评论数量(条)	名次	图书馆名称	评论数量(条)
1	上海图书馆	3 044	21	山西省图书馆	54
2	南京图书馆	1 033	22	宁夏图书馆	43
3	首都图书馆	922	23	福建省图书馆	41
4	广州图书馆	689	24	江西省图书馆	40
5	天津图书馆(文化中心分馆)	532	25	海南省图书馆	37
6	天津图书馆(复康路分馆)	498	26	安徽省图书馆	35
7	上海少年儿童图书馆	445	27	广西壮族自治区图书馆	35
8	广东省立中山图书馆(文明路总馆)	356	28	河南省图书馆	33
9	辽宁省图书馆	340	29	福建省少年儿童图书馆	28
10	浙江图书馆	307	30	天津少年儿童图书馆	24
11	湖北省图书馆新馆	303	31	河南省儿童图书馆	15
12	陕西省图书馆	290	32	甘肃省图书馆	9
13	四川省图书馆	266	33	贵州省图书馆	7
14	吉林省图书馆	128	34	南宁市图书馆	6
15	山东省图书馆	111	35	新疆自治区图书馆	4
16	黑龙江省图书馆	103	36	西藏图书馆	4
17	湖南图书馆	88	37	湖南省少年儿童图书馆	4
18	重庆市图书馆	68	38	内蒙古图书馆	2
19	云南省图书馆	65	39	重庆少年儿童图书馆	2
20	河北省图书馆	61			

表7 各因素之间相关性数值

	人均价格	书籍评分	环境评分	服务评分	好评数	中评数	差评数
人均价格	1.000 000	0.459 512	0.465 429	0.359 852	0.305 032	0.220 132	0.233 418
书籍评分	0.459 512	1.000 000	0.982 383	0.957 707	0.460 206	0.301 254	0.319 193
环境评分	0.465 429	0.982 383	1.000 000	0.961 812	0.428 305	0,270 717	0.299 829
服务评分	0.359 852	0.957 707	0.961 812	1.000 000	0.381 650	0.220 201	0.248 325
好评数	0.305 032	0.460 206	0.428 305	0.381 650	1.000 000	0.952 988	0.934 689
中评数	0.220 132	0.301 254	0.270 717	0.220 201	0.952 988	1.000 000	0.975 634
差评数	0.233 418	0.319 193	0.299 827	0.248 325	0.934 689	0.975 634	1.000 000

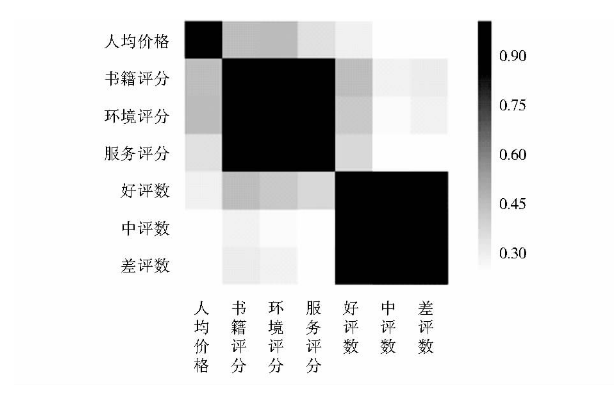


图4 各因素相关性热力图

2.4 用户评价情感分析

本文探究用户对公共图书馆形象的评价和情感倾向,从用户对图书馆的评论切入,对用户评论进行高频词统计和文本情感分析。大众点评用户对省级公共图书馆的评论共计10 072条评论。本研究对评论内容进行拆分词处理,剔除拆分词中的连接词、副词之类的词语,应用wordCloud对出现的词语进行展示制成词云图见图5,通过词云图可以看到不同单词和词组出现的频率高低差别。

对关键词依照词频由高到低进行排序,提取排名前20的高频关键词见表8。从这些高频关键词中可以发现用户的关注点主要在"图书""环境""借书""学习""氛围""阅览室""服务""安静""数字化"方面。此外,"藏书""人员""阅读""位置""周末""身份证""交通""外借""还书"等词出现次数也比较多。

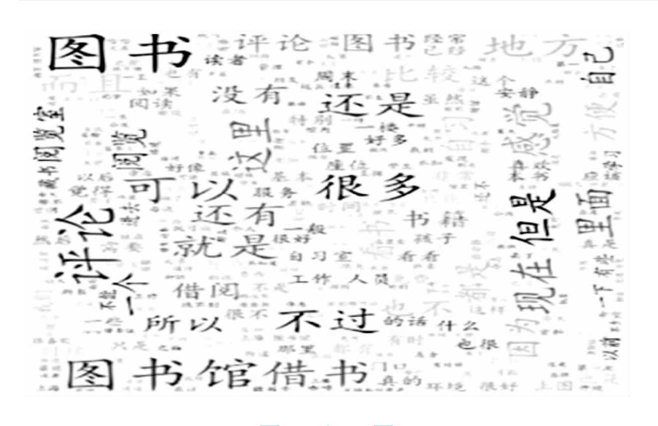


图5　词云图

表8　排名前20高频词

关键词	词频	关键词	词频
图书	4 767	免费	1 050
评论	4 469	书籍	1 007
图书馆	4 082	安静	980
环境	2 315	自习	974
借书	2 116	阅览室	951
方便	2 057	学习	974
看书、读书	1 846	工作	925
地方	1 829	服务	911
阅览	1 255	借阅	909
不错	1 105	喜欢	891
押金	1 054	数字化	868

本研究使用SnowNLP包进行中文文本的情感分析。SnowNLP的情感分析取值表达的是“这句话代表正面情感的概率”，其取值范围是(0,1)，情感值越接近于1即表达的情感越积极向上。通过观察样例评论数据以及其情感值(见表9)，发现大众点评网上用户对公共图书馆的评论大部分表现出积极态度。但是少量数据的观察，可能会造成结论上的偏颇。因此，为了避免这一现象，本研究使用mean()函数对情感分析结果数值进行处理，得到平均值是0.793，数值超过了0.7，说明整体上用户对公共图书馆的评价态度是正面的。接着，本研究对中位数进行计算，结果为0.996 8，发现中位数值不仅比平均值高，而且几乎接近1(完全正面)。这就意味着，大部分的评价表示非常满意，但存在着少部分异常点，显著拉低了平均值。大部分的满意评论可以用于未来公共馆建设的“扬长”，少部分异常点才是研究未来研究过程中应当尽力改进并及时弥补的欠缺点。根据发现，本研究对评论当中出现的优劣之处进行梳理，并提出相应对策建议以供后续研究参考。

表9　评论切分词表

频次	关键词	情感值
1 198	湖北省，图书馆，新馆，公正，路上，乘坐，566，817，站下，坐地铁，4号，于楚，河汉，2站，免费，可以，自由，阅读，书籍，如要，借书，回家，身份证，一百，押金，办理，一张，图书证，二百，外文书，借阅证，类型，本馆，通行，两类，中外文，读者，中文，发放，对象，中国，公民，与外，人士，愿意，遵守，规章制度，持有，证件，均可，申请，办证，手续，人工，填写，申请表，依据，自身，需要，确定，本人，有效，交纳，相应，即可，自助，根据，机系统，提示，每张，最多，3张，使用，范围，借阅，总量，4本，可借，少儿，图书，期刊，光碟，盲文，借期，30天，续借，声像，不可，1本，同时，3本，保持，开放，期限，默认，30年，超期，每天，每本，0.2元，挂失，丢失，办证处，工本费，5元，如在，找到，解除，继续，补办，不再，退还，原证，自动，作废，交相，还清，须由，在办，退押，出示，代办，2012年，12月，3日，以前，还需，携带，三个月，方可，后方，可再，位于，武汉市，武昌区，25号，占地，100.5亩，主体，建筑，地上，8层，地下，2层，面积，100 523，平方米，文献，藏量，可达，1000万，各类，阅览，坐席，6300个，接待，能力，10 000人，年数，资源，加工能力，500gb，在线，存储，150tb，所以有，机会，坐坐，感受一下，非常，值得，也非，常安，很多，努力，小伙伴，投入，学习，状态，会给，很大，鼓舞，，评论，湖北，省图书馆，馆站，下车，坐地，身份，外文，书刊，中外，规章，制度，张读，30，工本，2012，12，个月，武汉，武昌，25，100，平方，1 000，6 300，10 000，人次，加工，所以，有机，感受，小伙	0
7 425	高中，时候，经常，自习，地方，8层，格局，现在，已经，变了，还是，怀念，以前，座位，交通，比较，方便，很多，都到，不过，我们，都是，骑车来，一大早，还要，排队，图里，面机，各类，其实，挺多，还有，好多人，不知道，秘密，聚集地，续借，每次，回家，忘了，借完，先去，2层，电脑，借书，日期，一改，就不会，宝贝，过期，楼下，幸运，时间，虽然，东西，但是，放松，吃个，小美，哪怕，享受，吉野家，鸡肉，牛肉，或者，双拼，很有，怀旧，感觉，，评论，骑车，大早，好多，人不知，聚集，就不，吉野	2.22＊E^{-16}
5 981	之前，一直，徐汇，图书馆，因为，要找，论文，资料，去了，上海，据说，很多，可是，辗转，到了，那里，发现，没什么，第一，寄包，麻烦，也要，光通，检查，插卡，再放，第二，进一，都要，第三，多到，什么，地步，书明，明显，找来找去，找不到，心想，可能，还没，放上去，这时，阿姨，正在，旁边，几个，书架，放人，还回来，又分，3次，去找，始终，没有，再回，系统，依然，显示，火大，第四，办了，25，一年，外借，却被，告知，只能，借出，当天，阅览，那叫，分明，骗钱，还让，要么，1000，押金，那么多，在身，半天，借书，行程，重复，最后，借了，3本，小说，回家，看看，不然，真的，，评论，直去，图书，找来，不到，放上，还回，那么	4.44＊E^{-16}
6 287	最喜欢，地方，之一，想当年，一星期，2篇，论文，时候，在那里，常常，一天，见证，设备，设施，日益，先进，感受，浓重，静谧，阅读，氛围，泡上，香浓，咖啡，带上，手工，小饼干，拿上，自己，喜爱，书籍，这样，往往，阳光，包围，温暖，惬意，生活，现在，少了，活动，却要，经常，出入，好想，再去，安静，一整天，，评论，最喜，想当，星期，在那，天一，小饼，整天	0.999 999
7 115	去了，好多年，没事，就跑，那里，借书，现在，新馆，绝对，高端，大气，上档次，里边，还有，咖啡厅，还书，系统，都是，电子化，喜欢，图书馆，那个，便民，书店，有很，新书，推荐，一楼，展厅，经常，举办，展览，活动，欣赏，各国，文化，最方便，一点，就是，图书，别的，可以，就不，跑回，原来，，评论，好多，上档，咖啡，电子，馆里	1

（续表 9）

频次	关键词	情感值
7 805	真的，好棒，天天，都想，新的，四川省，图书馆，天府，广场，旁边，很漂亮，建筑，但是，也不，徒有，里面，更是，大气，内涵，进去，之后，安检，而且，稍微，大一，都要，只是，不明白，存完，为什么，还要，一次，负责任，了嘛，外借，就在，一楼，左右两，都有，门口，详细，信息，哪些，可以，在哪儿，找到，还有，很多，电脑，查到，这一点，方便，二楼，开始，就有，自习，看书，桌子，名字叫，星空，阅览室，超级，喜欢，星星，好美，名字，然后，书架，两边，开放，自己，坐在，自习室，其实，位置，挺多，人性化，插线板，阅读灯，还去，外文，本来，什么，期待，报刊，种类，和数，超出，想象，惊喜，更新，比较慢，最新，期刊，都是，几个月，一点，就是，时候，工作，人员，聊天，声音，很大，这样，阅读，氛围，就不，安逸，总体，来说，很棒，文化，服务设施，，评论，省图书馆，不明，负责，左右，在哪，这一，阅览，人性，插线，比较，几个，月前，服务，设施	1
3 505	外面，39 度，里面，凉快，第一次，保安，很热，老房子，木楼梯，干净，放着，沙发，喜欢，屋顶，复杂，石膏线，房子，肯定，故事，猫猫，一下子，人了，找到了，一本，杂志，一些，上海，小说，一篇，新乐，旧事，怀旧，文章，要是，咖啡，更好，远了，人家，儿童图书，有大，已经，很不错，总之，非常，这里，，评论，39，第一，老房，木楼，石膏，下子，找到，儿童，图书，很不，错了	1

3 研究结果与讨论

总体而言，本研究通过对我国 31 个省的 39 个省级公共图书馆在大众点评网上的公众评价进行现状调查、频数分析、影响因素偏度分析、评价因素相关性分析以及用户评价情感分析，得出研究结果如下：

3.1 均衡差异

研究数据显示，用户评价数据在地域和内容方面存在差异。

3.1.1 均衡地域分布差异 本研究对各区域内的省级公共图书馆按频数分析中的星级进行统计分布，结果显示华北、华东、东北 3 个地区的省级馆较之其他地区更受点评网用户喜欢。其中，华北和华东地区各 4 家，东北地区 3 家，华南和西南地区各 2 家，西北和华中地区各 1 家，呈现出一定的地区差异。此外，由偏度分析可知，我国省级公共图书馆的网上评论数值存在较大地区差异。偏度值最大的因素是读者在网上的评论数，其中，华北地区北京和天津评论数量多，内蒙古参与评论的人数最少；西北地区以陕西评论数量最高，甘肃、青海、新疆参与评论人数较少；华东地区以浙江、江苏和上海评论数量突出，安徽数值最低；东北地区数量分布较为均衡，以辽宁评论数量为首，其他几个省相差不大；西南以四川和重庆所在区域评论值较高；华南以广州评论数量最高；华中地区湖北评论数值最高，湖南次之。总体来说，地区分布的不平衡一定程度上反映出到馆人数上的差距，这一点在偏度分析中得以验证。因此，要均衡地区分布差异，应取长补短，建设区域特色资源，促进读者到馆阅读。

3.1.2 均衡评价内容差异 根据频数分析中的评分数值分布图（见图 3）可知，大众点评网对于省级公共馆的评分指标主要是书籍、环境、服务。其中在评分高的区间内，服务评分明显比其他两个指标的频数值低，说明省级馆在服务方面相对较弱。同时，服务与书籍和环境之间又有较强的相关性，提升图书馆的服务也会进一步促进用户对图书馆馆藏质量和环境的感知度。

3.2 增强对用户的吸引力

本研究相关性分析显示，用户评价数据整体的相关度高，数据具有良好的可信度。但是偏度分析结果显示，图书馆评论数的偏度存在很大差异，这表示应用大众点评网的用户评论的离散度高，部分区域的用户评价结果之间不具备可比性。这与我国区域经济和省级公共图书馆发展现状是相符的。因此，需要增强对用户的吸引力，尤其是偏远地区。从相关性分析结果来看，可以从馆藏资源、环境等方面的提升来促进对用户的吸引。一方面，提高用户的到访率，吸引更多的读者来图书馆享受阅读；另一方面，提高用户的参与度，给予网络评价的用户以一定的奖励，鼓励用户参与图书馆的评价。

3.3 关注用户情感倾向

用户的情感是了解读者的需求和期望的途径之一，有助于关注读者的兴趣点，从而提供有针对性的服务。通过高频词和词云图的分析，除了读者重点关注资源（图书、藏书）、服务（借书、还书、外借、阅读）、环境等内容，还应该关注安静、数字化、位置、身份证、交通等隐藏需求和期望；对于重点内容的把握，有助于图书馆有针对性地开展相关服务和活动。

关注读者的情感依赖。通过情感值的分析发现，读者对公共图书馆的评价是正面的。但是也有极少的异常值非常值得关注。以读者对馆员态度评价为例，在所获取的全部评论当中共有 26 条直接提到馆员服务态度的评论，其中差评 10 条，好评 8 条，不带个人主观色彩的中评 8 条。差评中多数提到馆员“个人素质差”“服务能力弱”“熟视无睹”“操作不熟”等词语。这些反馈直接影响用户对图书馆的满意度。因此，负向情感词有助于图书馆掌握用户的负面情绪，并及时应

对。

4 结语

本研究试图利用大众点评网数据对用户评论内容进行分析,应用频数分析、偏度分析、相关性分析和情感分析等方法揭示用户对我国省级公共图书馆的评价状况。本研究从用户评价的视角,以用户生成评价内容的数据为基础,揭示用户评价数据的客观性和倾向性,与以专家评价为主的公共图书馆评估方式相比,可以作为一种补充。一方面,评价数据可以同专家评价相互印证,另一方面,可作为用户满意度的补充。近几次公共图书馆评估都增加了读者满意度调查,以问卷调查的方式获取读者满意度,但是调查过程和被调查者易受到主观因素影响而达不到满意度调查的最初目的,基于用户网络评价数据相比而言更具客观性。此外,用户评论的情感分析还可以对读者的使用倾向进行预判,可以对正向情感进行强化,对负向情感进行控制。

本研究仍然存在一定局限性。首先,大众点评网的读者对公共图书馆评价数据的局限。由于大众点评网数据结构的固化,只能解释书籍评分、服务评分、环境评分、人均价格、好评、中评和差评,而无法揭示更多评价内容。其次,本研究相关性的结果只能显示相关关系,并不能解释原因,需要进一步深入研究。第三,本研究用户评论的情感分析仅仅对情感值进行赋值获取整体情感评价,未对具体情感倾向进行赋值,具体情感倾向需要进一步的后续研究进行揭示。

参考文献:

[1] 宋运郊.美国公共图书馆系统的最低标准[J].黑龙江图书馆,1984(2):63－67.
[2] 施士宇.尼日利亚的公共图书馆[J].图书馆杂志,1984(2):71－72.
[3] 王丽萍.县图书馆评估初探[J]. 图书馆学刊,1987(5):5－7.
[4] 李东来,王孝龙,黄丽华.图书馆评估研究纲要[J].图书馆学刊,1987(2):1－7.
[5] 徐淑兰,张雁翎. 市级公共图书馆的评价[摘要][J].图书馆建设,1988(1):16－17.
[6] 黄明,易绛.试论公共图书馆藏书的评估[J].图书馆工作与研究,1988(3):50－55.
[7] 徐隆沽.图书馆评估指标体系的建立及其科学性[J].图书馆工作与研究,1993(1):53－55.
[8] 蒋金葆.关于确立图书馆评估指标体系基本原则的思考[J].图书馆理论与实践, 1995(1):24－26.
[9] 彭绮文.图书馆评估的指标体系[J].中山大学学报(社会科学版),1996(2):135－138.
[10] 周礼智.图书馆专业人员绩效评估浅探[J].图书馆学刊,1991(1):22－25.
[11] 张晓林.藏书评价研究[J].四川图书馆学报,1996(3):41－44.
[12] 王学东,薛朗.图书馆藏书评价研究[J].图书情报知识, 1994(3):19－22.
[13] 张琳.论藏书效用评价[J].图书情报工作,1998(8):26－28.
[14] 付立宏,李祁平.近年西方图书馆参考服务评价研究述评[J].江苏图书馆报,1994(5):19－21.
[15] 郭兴超.现代图书馆的信息服务及其评价[J].图书馆理论与实践,1996(2):9－12.
[16] 夏军.简述国内图书馆效益评价的理论与方法[J].图书馆工作与研究,1995(4):23－25.
[17] 侯丽京.图书馆效益及评价方法[J].中南民族学院学报(自然科学版),1998(4):90－93.
[18] 龙滔.公共图书馆图书采购投资的智能模糊综合评价系统[J].图书馆理论与实践,2000(4):20－21.
[19] 许玲,王利,周从军.面向全面质量管理——用 SERVQUAL 评价图书馆服务质量[J].图书情报工作,2001,45(5):76－79.
[20] 任馨,章云兰,金中仁.图书馆服务质量评价新模式——从 SERVQUAL 到 LibQUAL＋[J].情报杂志,2002(7):47－48.
[21] 颜爱民,胡丕志,廖伟.平衡计分卡思想和 DHP 方法在公共图书馆绩效评估中的应用[J].图书馆,2005(3):77－80.
[22] 李国新."十二五"时期公共图书馆事业的发展机遇[J].图书馆建设,2011(10):2－6,11.
[23] 李丹,申晓娟,王秀香,等.新起点 新视野 新任务——第五次全国公共图书馆(成人馆部分)评估定级标准解读[J].中国图书馆学报,2013,39(2):4－17.
[24] 吴正荆,袁艺,郝森森.第五次公共图书馆评估结果分析[J].图书馆建设,2014(12):88－91,95.
[25] 柯平,宫平.全国公共图书馆第六次评估的意义和特点[J].图书馆建设,2016(12):4－7,14.
[26] 李仁德,张勤,王细荣.以用户满意为中心的图书馆服务评价模式构建[J].图书馆论坛,2012,32(1):113－116,19.
[27] 施国洪,陈洁,沙振江,等.基于 LibQUAL＋TM 的我国公共图书馆服务质量评价模型构建研究[J].图书馆论坛,2014,34(2):7－14.
[28] 李建霞.图书馆绩效评价研究综述[J].图书情报知识,2011(5):42－52.
[29] 赵治卿.平衡记分卡理论在图书馆绩效评估中的应用初探[J].图书馆工作与研究,2014(3):43－46.
[30] 冯琳,高波.美国公共图书馆绩效评估理论研究[J].图书馆建设,2012(3):22－26.
[31] 苏福,柯平.公共图书馆评估的关键指标探讨——以省级公共图书馆为例[J].图书馆建设,2016(12):15－20.
[32] 百度百科.大众点评网[EB/OL].https://baike.baidu.com/item/大众点评网/3359289? fr＝Aladdin.
[33] 赵鹏,骆德汉,梅领亮.一种基于 JSON 的 Android 终端远程获取解析数据的方法[J].微型机与应用,2017,36(21):57－60.

[34] 王亨. 标准差 标准误 偏度系数和峰度系数的作用[EB/OL]. [2018-09-13]. https://blog.csdn.net/wzgl__wh/article/details/52080688.

作者贡献说明:

张文亮:提出研究思路、内容和框架,修改论文;
付枭男:负责文献调研、数据获取、论文撰写与修改;
王彦芳:负责数据获取、数据分析、论文撰写与修改。

Research on Evaluation of Provincial Public Library
——Taking Dianping.com Data as an Example

Zhang Wenliang Fu Xiaonan Wang Yanfang
School of Goverment, Liaoning Normal University, Dalian 116021

Abstract: [**Purpose/significance**] The purpose of this research is to explore the status of evaluation of provincial public libraries in China, and to compare the construction differences of 39 provincial public libraries in 31 provinces in 7 regions of China, including East, Central, South, North, Northwest, Southwest and Northeast area, so as to provide suggestions for the future construction of public libraries. [**Method/process**] Based on the Basic Data of the Development of Chinese Public Libraries, the Yearbook of Chinese Library 2016, and the website of the National Bureau of Statistics of the People's Republic of China, it obtains the objective evaluation data of public libraries and determines some data acquisition dimensions. Using Python web crawler technology to obtain the assignment score and user comments on public libraries in China, this paper comprehensively analyzes the relationship between the influencing factors of public library service evaluation in China and puts forward some constructive suggestions. [**Result/conclusion**] Based on the data analysis, the paper conclusions are drawn: equalize differences involving the difference of regional and the content of evaluation; enhance the attractiveness of users, especially in remote areas; pay attention to the user's emotion tendency, focus on the hot words of readers ' attention, and at the same time pay attention to eliminate the influence of negative emotion

Keywords: library evaluation Python evaluation analysis big data